1992

WISC-R and WISC-III aptitude scores: How comparable are they?

Timothy John Fisher
University of Northern Iowa

WISC-R AND WISC-III APTITUDE SCORES: HOW COMPARABLE ARE THEY?

An Abstract of a Thesis

Submitted

In Partial Fulfillment

of the Requirements for the Degree

Specialist in Education

Timothy John Fisher

University of Northern Iowa

July 1992

ABSTRACT

Fisher, Timothy John. Specialist in Education, University of Northern Iowa, July 1992. WISC-R AND WISC-III APTITUDE SCORES: HOW COMPARABLE ARE THEY?

This study examined the comparability of the WISC-R and WISC-III subtest and scale scores. The Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility indices were also examined for comparability.

The researcher administered the WISC-R and WISC-III in a counterbalanced design to 20 rural, Caucasian, midwestern children ranging in age from 8 to 12 years old. On both the WISC-R and WISC-III, means and standard deviations were computed for the subtest scaled scores, scaled index scores, and Verbal, Performance, and Full Scale IQs. Correlation coefficients were obtained for the WISC-R and WISC-III subtests, indices, and scales with two tailed dependent *t* tests computed to determine comparability.

Results from *t* test analysis revealed significantly higher WISC-R subtest scores for the Vocabulary, Comprehension, and Block Design subtests and a significantly higher WISC-III Information subtest score. A significantly higher score was also found for the WISC-R Verbal Comprehension index and all three WISC-R IQ scales. The difference in IQ scale scores was similar to the results reported in the only published study comparing the two instruments.

Correlation coefficients indicated a significant correlation for all subtests, indices, and scales with the exception of the Comprehension and Picture Arrangement subtests, and the Perceptual Organization index.

Historically, various research comparing the differences in IQ scores of the same child on the older and newer versions of the same intelligence test have been conducted. This study appears to coincide with this research in that the group IQ scores of the subjects are lower on the newer, revised version of the test. It is of great importance that teachers, parents, and psychologists be aware that IQ scores on the WISC-R and WISC-III appear to be somewhat discrepant. Because the WISC-III is still very new and psychologists will inevitably be changing from the WISC-R to the WISC-III, additional investigations are warranted for further study.

This Study by: Timothy John Fisher

Entitled: WISC-R and WISC-III Aptitude Scores: How Comparable Are They?

has been approved as meeting the thesis requirement for the Degree of Specialist in Education

7/13/92
Date — Dr. Ralph Scott, Chair, Thesis Committee

7-13-92
Date — Dr. Barry Wilson, Thesis Committee Member

7-13-92
Date — Dr. Augustine Osman, Thesis Committee Member

8-24-92
Date — Dr. John W. Somervill, Dean, Graduate College

ACKNOWLEDGMENTS

First of all, I would like to express my appreciation to the members of my thesis committee, Dr. Barry Wilson and Dr. Augustine Osman, for their time, expertise, helpful suggestions, and support. A very special thank you to my thesis chairperson, Dr. Ralph Scott, for all of his assistance, expertise, enormous supply of patience, and his wonderful Scandinavian sense of humor.

I would also like to acknowledge the past and present graduate students who have touched my life during the past three years. I could not have survived and enjoyed this experience without your help. Your friendship will never be forgotten. I want especially to thank my colleagues, Andy Smith, Dave Wood, Sue Schultz, Laura Bealles (Be), Laura Clark, (Bo), and Karen Annis for helping to actually make graduate school fun and memorable from beginning to end.

Most importantly, I want to express my deepest gratitude and appreciation to my family--my son, Cody, and my new daughter, Kaylee, for their understanding, and especially to my wife, Sue, for all of her support and understanding through both good times and difficult times. Without her continued support and sacrifices, I would not be obtaining this degree.

WISC-R AND WISC-III APTITUDE SCORES: HOW COMPARABLE ARE THEY?

A Thesis

Submitted

In Partial Fulfillment

of the Requirements for the Degree

Specialist in Education

Timothy John Fisher

University of Northern Iowa

July 1992

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Currently the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974), an assessment tool used widely in assessing the intellectual functioning of children, is being replaced by another updated version, the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991). With the publication of this revision, concerns have been raised regarding the comparability of IQ scores obtained on the WISC-R and WISC-III. The concerns entail questions such as: Is there a significant difference in subtest scores on the WISC-R compared to the WISC-III?; Does a significant difference occur between WISC-R and WISC-III IQ scale scores?; Are the two instruments significantly correlated? Similar concerns were noted by numerous researchers (Covet, 1976; Gironda, 1977; Paal, Hesterly, &, Wepfer, 1979; Mishra & Brown, 1983; Lippold & Claiborn, 1983) when a revision of a Wechsler instrument was compared with its predecessor. Sattler (1982, p. 150-151), noted that lower IQ scores were obtained on the WISC-R when compared to the original Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949). He stated that in retest situations, changes within these limits on the WISC-R should be viewed as a reflection of changes in the test norms and not as a decrease in the child's level of intellectual functioning.

Although the limitations of IQ scores themselves are generally acknowledged, scores obtained on IQ tests often play a significant role in the determination of the educational placement and programs determined most appropriate for children. For example, an IQ score on a standardized test instrument greater than one standard deviation below the mean for that instrument is one of the qualifying factors for classification as a mental disability (Iowa Department of Education, 1990). Also, interventions are often based on apparent strengths and weaknesses implied from achieved intelligence test scores (Sattler, 1988, p. 822-829). This study provides empirical information concerning the comparability of scores on the two instruments.

Statement of the Problem

The primary purpose of this study is to examine the comparability of WISC-R and WISC-III subtests, indices, and scale scores for a sample of children between the ages of 8 and 12 years old. Four hypotheses will be examined employing statistical probability at the .05 level, with no specific directionality hypothesized.

The first hypothesis is there is no significant difference in any of the WISC-R and WISC-III Verbal Scale and Performance Scale subtest scores.

The second hypothesis is there is no significant difference in WISC-R and WISC-III scores on the Verbal

Comprehension, Perceptual Organization, and Freedom from Distractibility indices.

The third hypothesis is there is no significant difference in WISC-R and WISC-III scores on the Verbal, Performance, and Full Scales.

The final hypothesis is there is no significant correlation in WISC-R and WISC-III subtests, indices, and scales.

Significance of the Study

If the scores obtained on the WISC-R and WISC-III significantly differ, future educational programming and the incidence of children identified for services could be greatly affected. Presently, in some school districts, specific criterion scores legally qualify students for special programming. In addition to other criteria, some children qualify for special programs because of only one or two IQ points. Separately, if subtest, index, or scale scores are significantly discrepant on the WISC-R and WISC-III, then clinical conclusions based on such discrepancies can lead to erroneous or inconsistent educational strategies, thus undermining credibility of aptitude tests. For example, when being originally evaluated for educational assistance a hypothetical 10 year old child obtained a Verbal Scale IQ score of 85 and Performance Scale IQ score of 100 on the WISC-R. The difference of 15 points between the two scales is considered significant and reflects a high probability

that levels in verbal and nonverbal intellectual functioning are different and a language deficit may exist (Sattler, 1988, p. 167). As a 12 year old, the same hypothetical child was re-evaluated using the WISC-III. The obtained Verbal Scale and Performance Scale IQ scores were 93 and 85, respectively, with the differences not considered significant. A concern arises whether the change in scale score differences are attributable to test instrument differences or other factors.

Assumptions

1. The WISC-R and WISC-III are considered to be valid measures of intelligence.

2. Comparison of each child with himself or herself, based on scores obtained (by each child) on the WISC-R and WISC-III, has minimized norming problems.

Limitations of the Study

1. The subjects of the design are rural, Caucasian, midwestern children; they do not constitute a sample representative of the population as a whole.

2. The sample is relatively small.

3. The researcher administered all the tests and scored all the protocols which is a potential source for bias.

4. Previous research comparing the WISC-R and WISC-III is limited to one study and that is in the WISC-III Manual (Wechsler, 1991, p. 197-199).

5. This comparison of the WISC-R and WISC-III may serve clinicians in better understanding the intellectual functioning of a specific child but not serve as a sole measure.

Definition of Terms

1. Intelligence--Snyderman and Rothman (1986) report that there is a high degree of consensus among experts in the fields of psychology, sociology, and genetics as to what they believe are important elements of intelligence. With near unanimity (96% agreement among those who answered the survey regarding the definition of intelligence), the experts described intelligence as consisting of abstract thinking or reasoning; the capacity to acquire knowledge; and problem solving ability. A majority of respondents (60% to 80%) checked the following as characteristics of intelligence; adaptation to one's environment, creativity, general knowledge, linguistic competence, memory and mental speed.

2. Reliability--This refers to the degree of consistency with which a measuring instrument measures whatever it is measuring. A test is reliable to the extent that the scores made by an individual remain nearly the same in repeated measurements (Anastasi, 1976, p. 27).

3. Factor based analysis--This is a mathematical procedure used to analyze the intercorrelations of a group of tests that have been administered to a large number of individuals. Factor analysis is based on the assumption that

intercorrelations can be accounted for by some underlying set of (unobservable) factors that are fewer in number than the tests themselves. A major purpose of factor based analysis is to simplify the description of behavior by reducing the number of variables to the smallest possible number (Sattler, 1988, p. 31-32).

4. Verbal Comprehension (VC)--This factor index is comprised of the Information, Similarities, Vocabulary, and Comprehension subtests of the Verbal Scale. It is comprised of the same subtests on both the WISC-R and WISC-III (Wechsler, 1991, p. 187). The VC factor score measures verbal knowledge and understanding obtained informally and through formal education. It reflects the application of verbal skills to new situations (Sattler, 1988, p. 130).

5. Perceptual Comprehension (PO)--This factor index consists of the Picture Completion, Picture Arrangement, Block Design, and Object Assembly subtests of the Performance Scale and is comprised of the same subtests on both instruments (Wechsler, 1991, p. 187). It reflects the ability to interpret and organize visually perceived material within a time limit (Sattler, 1988, p. 131).

6. Freedom from Distractibility (FFD)--This factor index consists of the Arithmetic and Digit Span subtests of the Verbal Scale (Wechsler, 1991, p. 187). This factor score measures the ability to attend or concentrate, but also may involve numerical ability. Short-term memory may be an

important component of the FFD factor index (Sattler, 1988, p. 131) On the WISC-R the Coding subtest also loads on this factor (Sattler, 1988, p. 131). On the WISC-III, however, the Coding subtest combines with a newly introduced WISC-III subtest, Symbol Search, to form a new factor (Wechsler, 1991, p. 187). To maintain consistency in this study, only the Arithmetic and Digit Span scaled scores were used to compute the FFD index.

7. Verbal Scale--This scale is composed of the Information, Similarities, Arithmetic, Vocabulary, and Comprehension subtests and a supplementary subtest, Digit Span (Wechsler, 1991, p. 5). The Verbal Scale indicates verbal comprehension, application of verbal skills and information to the solution of new problems, verbal ability, ability to process verbal information, and ability to think with words (Sattler, 1988, p. 856).

8. Performance Scale--This scale is composed of Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding subtests; and Mazes, a supplementary subtest (Wechsler, 1974, p. 8). The WISC-III has added an additional supplementary Performance Scale subtest, Symbol Search (Wechsler, 1991, p. 5). The Performance Scale is associated with perceptual organization, ability to think in terms of visual images and to manipulate them with fluency and flexibility; ability to interpret or organize visually perceived material against a time limit; nonverbal ability;

and ability to form relatively abstract concepts and relationships without the use of words (Sattler, 1988, p. 856).

CHAPTER 2

REVIEW OF THE LITERATURE

The literature reviewed in this chapter is organized into three sections. The first section highlights the historical trends of IQ scores and inferences that can be made from these trends. Section two includes a review of empirical data concerning the comparability of the WISC-R and WISC-III. The third and final section of this chapter presents a representative sample of research comparing revised Wechsler instruments with their predecessors using comparison of means and correlation coefficients in data analysis.

Historical Trends and Inferences

A concern surrounding the historical development of intelligence tests has raised many questions regarding the comparability of scores obtained on the WISC-R and WISC-III. Since the onset of tests devised to measure intellectual abilities, the stability of IQ scores has been debated. One central controversy involves whether IQ scores have increased or decreased over time. Flynn (1984, 1987) and Kaufman (1990) have reported that the average IQ scores of Americans have steadily risen since the 1930s. Every standardization sample for the Stanford-Binet and Wechsler instruments from 1932 to 1978 established higher norm standards than its predecessor. An effort was made by Flynn (1984) to locate every study in which two or more tests were administered to

the same group of subjects. To promote accuracy, every score was translated into so many standard deviations above or below the mean, assigning the mean and standard deviation values of 100 and 15, respectively. The total gain amounts to a rise in mean IQ of 13.8 points on the Wechsler instruments (Flynn, 1984). Consistent with this trend the more recent forms of IQ tests yield lower IQ scores of approximately five points (Flynn, 1984). Sattler (1982) viewed these trends as a reflection of the updating of norms for the revised instrument and not a decrease in an individual child's intellectual functioning. According to the WISC-III Manual differences between WISC-R and WISC-III IQ scores are expected and similar to those reported between previous revisions of Wechsler instruments (Wechsler, 1991, p. 197).

These trends infer that children must, in effect, score higher on the newer forms of intellectual tests in order to achieve equivalent scores as those obtained on an earlier version of the same assessment tool. Should this be true, program planning for thousands of educationally evaluated children could be directly affected, since classification is, at least, partially based on IQ scores. The Iowa Rules of Special Education state that

> "mental disability" is the inclusive term denoting significant deficits in adaptive behavior and subaverage general intelligence. For educational purposes, adaptive behavior refers to the individual's

> effectiveness in meeting the demands of one's environment and subaverage general intellectual functioning such as evidenced by performance greater than one standard deviation below the mean on a reliable individual test of general intelligence valid for the individual pupil. (Iowa Department of Education, 1990, p. 5)

For example, given a hypothetical student in the state of Iowa who obtained a Full Scale IQ score of 86 on the WISC-R, this child would not be considered mentally disabled. However, that same child receiving a Full Scale IQ score of 81 using the lower scoring WISC-III would be considered mentally disabled and therefore eligible for special education services.

Empirical Data Comparing the WISC-R and WISC-III

Unfortunately little empirical data exist concerning the comparability of the WISC-R and WISC-III. In an effort to obtain further empirical data, the author conducted both an ERIC computer search as well as a manual search of the most recent bound journal abstracts. This research yielded only one such study, and this was reported in the WISC-III Manual (Wechsler, 1991, p. 197-199). Here, reference is made to a study comparing the WISC-R and WISC-III to establish convergent and discriminant validity. The two tests were administered to a sample of 206 children 6-16 years old (median chronological age = 11); (Wechsler, 1991, p. 197). Half of the subjects were alternately assigned to receive the WISC-R first and the other half of the subjects received the WISC-III first (counterbalanced design). A comparison of the

WISC-R Full Scale mean IQ scores revealed that the WISC-III Full Scale IQ scores were approximately five points lower than the WISC-R Full Scale IQ scores, and that the WISC-III Verbal IQ and Performance IQ scores were approximately two and seven points less than the corresponding WISC-R IQ scores, respectively.

Separately, an unpublished report (Fisher, 1992) compared the WISC-R and WISC-III scores of 10 subjects and reported the Verbal, Performance, and Full Scale IQ scores were higher on the WISC-III by one to five points. However, the administration of the tests was not counterbalanced. Further, some test administrators were students still participating in a school psychology training program and had not yet successfully met course requirements on the administration of Wechsler instruments.

Review of Literature Comparing Other Wechsler Instruments

Historically, various versions of the Wechsler instruments have been examined as revised instruments were developed. A study by Davis (1977) used matched pair comparison of WISC and WISC-R scores to determine if IQ scores are significantly affected by the order in which the two tests are given. The author stressed counterbalancing when comparing two similar instruments.

Previous research has compared scores of the WISC with the then updated version, the WISC-R. Catron and Catron (1977) assessed WISC and WISC-R IQ scores from 62 educable

mentally retarded children in a counterbalanced design. All three WISC IQ scores (Verbal, Performance, and Full Scale) were significantly higher than the comparable WISC-R IQ scores. The average WISC Verbal IQ was more than five points higher than the average WISC-R Verbal IQ. The Performance IQ on the WISC was more than six points higher than the WISC-R IQ, and the Full Scale IQ's differed by approximately 5.5 points, with the WISC being higher. Further analysis which explored subtest differences revealed that differences in the Verbal IQ scores was largely due to the lower WISC-R scores on three of the five Verbal subtests: Information, Arithmetic, and Similarities. Differences in the Performance IQs were largely attributed to the marked differences in the Block Design and Coding subtests. The authors explained the lower WISC-R scores on those two subtests as resulting from more stringent norms on the WISC-R. These results were similar to the results from earlier published reports by Schwarting (1976) and Hamm, Wheeler, McCallum, Herrin, Hunter, and Catoe (1976) on the comparability of the WISC and WISC-R.

Various studies have assessed other Wechsler instruments and have included analysis of data by using a comparison of means and correlation coefficients. Quereshi and McIntire (1984) administered the WISC, WISC-R, and the Wechsler Preschool and Primary Scales of Intelligence (WPPSI; Wechsler, 1967) in counterbalanced design to 72 randomly

selected elementary school children in order to determine their comparability by testing the equality of (a) means; (b) variances; (c) reliability coefficients; and (d) validity coefficients based on scaled scores and IQs. Results indicated that the Verbal IQs were comparable, but the Performance and Full Scale IQs were not. The difference in Verbal mean IQs was less than one point with the WISC-R being higher. Conversely, the Performance mean IQs yielded a difference of approximately five points with the WISC score being higher than the WISC-R. The difference in Full Scale mean IQs between the WISC and WISC-R was approximately two points with the WISC score being higher. The Performance Scale and Full Scale results are similar to previous research regarding the difference in mean IQ scores on the Wechsler revision (Flynn, 1984; Catron & Catron, 1977). Among the subtests, only Comprehension, Arithmetic, Picture Completion, and Mazes were significantly correlated. Possible sources of apparent discrepancies among other subtests were discussed because some subtests that appeared to be comparable turned out not to be so, while some others that initially appeared to be different from each other emerged as comparable subtests.

Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) and Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) IQs and subtests were contrasted by Mishra and Brown (1983). Here the two scales were administered to a

sample of 88 adults in a counterbalanced order and results examined by correlational and repeated measure three-way analysis of variance procedures. Correlations between the two IQs for Verbal, Performance, and Full Scales were .83, .68, and .76, respectively. The average correlation coefficient was .77 for the six Verbal subtests and .64 for all the Performance subtests suggesting significant similarity between the two tests. Among all the Verbal subtests, the Vocabulary subtest of the two scales correlated the highest (r = .85), whereas the correlation between the Block Design subtest of the WAIS and WAIS-R was highest among Performance subtests. Comparison of mean IQs revealed that the WAIS Verbal, Performance, and Full Scale IQs were approximately five to six points higher than the corresponding IQs on the WAIS-R. These results were consistent with data reported by earlier studies that compared revisions of the Wechsler and Stanford-Binet scales with older tests (Flynn, 1984).

Lippold and Claiborn (1983) administered the WAIS and WAIS-R in counterbalanced order to thirty veterans referred for neuropsychological evaluation. A procedure was used that avoided the repetition of identical items. All correlations between the two forms were significant and ranged from .76 to .99. The IQ scores and all subtest scores were compared and found to be significantly different, with the WAIS scores being approximately seven to nine points higher than the

WAIS-R scores. Comparison of the two test forms found them to be reliable but not equivalent, and the difference in test scores is considered large enough to be important to the clinician.

In summary, most of the research presented (Flynn, 1984; Catron & Catron, 1977; Mishra & Brown, 1983; Lippold & Claiborn, 1983) has empirically shown that significantly higher IQ scale and subtest scores are obtained when comparing the revision of a Wechsler intellectual testing instrument with its predecessor. The high individual subtest correlations indicate consistency is maintained with subsequent Wechsler test revisions.

CHAPTER 3

DESIGN OF THE STUDY

Subjects

The subjects were 20 rural, Caucasian, midwestern children ranging in age from 8 to 12 years old. Subjects consisted of 10 boys with a mean age of 10.7 years and 10 girls with a mean age of 9.5 years. All subjects were Iowa residents and students at a rural, consolidated elementary school and all were tested by the researcher.

Instruments

The standard versions of the WISC-R and the WISC-III consist of ten subtests which are categorized as either Verbal or Performance Scale subtests. The Verbal and Performance Scales also consist of supplemental subtests which are Digit Span (Verbal) and Mazes (Performance). The WISC-III has an additional Performance Scale supplementary subtest, Symbol Search. Both instruments provide three separate IQs: a Verbal Scale IQ, a Performance Scale IQ, and a Full Scale IQ. All three are Deviation IQs obtained by comparing the subjects' scores with those earned by a representative sample of age peers. Because Deviation IQs are standard scores, at each age level each of the three IQs has the same mean of 100 and standard deviation of 15 (Sattler, 1988, p. 121). Kaufman's (1975, 1979) extensive analyses of the standardization data of the WISC-R clearly identified two major factors, VC and PO, and introduced the

notion of a third factor, FFD. In addition to the Verbal, Performance, and Full Scale IQ scores, these three factor based index scores can be calculated and converted over to comparable IQ scores. Whereas on the WISC-R FFD is normally associated with three subtests, on the WISC-III it is only cited as comprised of two. Therefore in the present study FFD will consist of the Arithmetic and Digit Span subtests. The WISC-III has added a fourth index, Processing Speed, consisting of Coding and Symbol Search, which may be calculated (Wechsler, 1991). However, no comparable subtest to Symbol Search is included in the WISC-R so therefore this index could not be employed in the present design. The subtests on both testing instruments are administered in alternating order beginning with Information, a Verbal Scale subtest, on the WISC-R. In contrast, the WISC-III begins with the administration of Picture Completion, a Performance Scale subtest.

Verbal Scale

Following is a brief description of the abilities and traits assessed by the Verbal Scale subtests of the WISC-R and WISC-III:

1. Information--assesses the child's general range of information, alertness to the environment, social or cultural background, and attitudes toward school and school-like tasks. Successful performance on this subtest requires memory for habitual, over-learned responses.

2. Vocabulary--taps a variety of cognition-related factors including learning ability, fund of information, richness of ideas, memory, concept formation, and language development that may be closely related to children's experiences and educational environment.

3. Arithmetic--requires the child to follow verbal directions, concentrate on selected parts of questions, and use numerical operations.

4. Similarities--assesses the ability to place objects and events together in a meaningful group or groups.

5. Comprehension--success on this subtest depends, in part, on possession of practical information plus an ability to draw on past experiences in reaching solutions. Responses may reflect the child's knowledge of conventional standards of behavior, extensiveness of cultural opportunities, and level of development of conscience or moral sense.

6. Digit Span--measures short-term auditory memory and attention.

Performance Scale

Following is a brief description of the abilities and traits assessed by the Performance Scale subtests of the WISC-R and WISC-III:

1. Picture Completion--assesses the ability to differentiate essential from nonessential details. This subtest requires concentration, reasoning, visual organization, and long-term visual memory.

2. Picture Arrangement--measures children's ability to comprehend and evaluate a total situation. Planning ability, anticipation, visual organization, and temporal sequencing are involved.

3. Block Design--involves the ability to perceive and analyze forms by breaking down a whole into its component parts and then assembling the components into the identical design, a process referred to as analysis and synthesis. The subtest requires perceptual organization, spatial visualization, and abstract conceptualization.

4. Object Assembly--requires visual-motor coordination, with motor activity guided by visual perception and sensorimotor feedback. It is also a test of visual organizational ability.

5. Coding--taps the ability to learn an unfamiliar task and involves speed and accuracy of visual-motor coordination, attentional skills, short-term memory, cognitive flexibility, and possibly, motivation.

Procedures for Test Administration

The researcher met with school officials of a rural, consolidated school in Iowa and explained the purpose of the study. A sample interest form briefly describing the study was given to the school officials. A cover letter from the school and the interest form were sent home with all third through sixth grade students. Of the 200 forms sent out, thirty were returned to the school with parental signature of

interest (13% response). After receiving the signed interest forms from the school, the researcher contacted the respondents by phone until 10 male subjects and 10 female subjects were identified for inclusion in the study. The purpose of the study was repeated to the parents during the initial phone contact. Any questions regarding their child's involvement in the study were addressed at that time and the initial appointment for test administration was scheduled. A suggestion by the researcher for testing the child at their residence to promote a comfortable, non-threatening atmosphere was agreed upon by all participants except one. The one subject not tested at their residence was tested at the school after dismissal.

In every case test administration was in a quiet room which was free of distractions. Before beginning administration of the first test the purpose of the study was repeated to the parent. A consent form was obtained with the child's name and parent signature as required by the Human Subjects Review Committee of the University of Northern Iowa in Cedar Falls. The parent received a copy of the signed consent form. The appointment for administering the second instrument was scheduled by phone or at the time of the initial test administration. The researcher administered the WISC-R and WISC-III in a counterbalanced design. Both tests were administered within 21 days of each other with a test administration mean difference of 17 days.

Procedures for Data Analysis

In this study both the WISC-R and WISC-III standard 10 subtests plus a supplemental Verbal subtest, Digit Span, were administered. The VC, PC, and FFD index scores, as well as the Verbal, Performance, and Full Scale IQ scores were calculated. On both the WISC-R and WISC-III, means and standard deviations were computed for the subtest scaled scores, scaled index scores, and Verbal, Performance, and Full Scale IQ scores. IQ equivalents for the indices were obtained for easier reader comparability. The WISC-R Manual does not provide IQ equivalents for the VC, PO, and FFD index scores. Therefore, Tables A.4-A.7 from the WISC-III Manual (Wechsler, 1991, p. 251-257) were used to convert the WISC-R and WISC-III index raw scores to IQ equivalent scores. Correlation coefficients were obtained for the WISC-R and WISC-III subtests, indices, and scales. Two tailed dependent *t* tests were also computed to determine the comparability of the subtests, indices, and scales.

CHAPTER 4

RESULTS AND DISCUSSION

This study was designed to examine the comparability of the WISC-R and WISC-III scores for a sample of 8 to 12 year old students. While the issue of comparability of previous Wechsler instruments has been addressed in the literature, only one published study to date has compared IQ profiles of the WISC-R with profiles on the most recent revision, the WISC-III. Explored in this study was the comparability of the subtests, indices, and scales of the WISC-R and WISC-III.

Results

Four hypotheses were tested for significance. The first hypothesis stated there is no significant difference in any of the WISC-R and WISC-III subtest scores.

This hypothesis was tested for each pair of subtests using a *t* test for dependent means and was tested for statistical significance. As noted on Table 1, WISC-R scores on the Information, Vocabulary, Comprehension, and Block Design subtests were significantly different than comparable subtest scores on the WISC-III. The largest single discrepancy appeared on Block Design where the mean subtest difference was 1.8 scaled score which is significant at the .001 level. Thus the null hypothesis of no significant difference in any of the subtest scores was rejected for the Information, Vocabulary, Comprehension, and Block Design subtests and accepted for all other subtests.

Table 1

WISC-R and WISC-III Standard Scores, Means, and Standard Deviations by Subtest, Scale, and Index

	Test						
	WISC-R		WISC-III			2-tailed	
	Mean	SD	Mean	SD	*t*	prob.	*r*
Verbal subtests							
Information	9.6	3.3	10.6	3.8	-2.8	.011*	.92**
Similarities	11.4	3.2	10.3	3.4	1.9	.073	.69**
Arithmetic	10.8	3.6	10.4	3.9	.8	.464	.80**
Vocabulary	9.8	2.3	8.6	3.5	2.4	.030*	.73**
Comprehension	11.4	2.7	9.6	3.0	2.4	.025*	.33
Digit Span	10.2	3.3	11.0	3.0	-1.8	.089	.75**
Performance subtests							
Picture Completion	11.2	3.1	10.3	2.8	1.4	.189	.52*
Picture Arrangement	12.6	3.7	10.8	4.1	1.9	.078	.22
Block Design	12.6	3.7	10.8	3.7	4.4	.001***	.87**
Object Assembly	12.0	.7	11.0	.8	1.4	.177	.52*
Coding	11.0	3.9	10.3	3.9	1.0	.347	.65**
Scales							
Verbal IQ	104.5	15.0	100.2	17.5	2.2	.041*	.84**
Performance IQ	112.8	15.3	104.3	15.7	2.5	.022*	.54*
Full Scale IQ	110.2	16.2	102.1	16.1	3.5	.002**	.79**
Indices							
VC	106.2	13.1	99.6	16.5	3.4	.003**	.85**
PO	113.2	13.1	110.2	29.6	.5	.652	.25
FFD	103.4	18.0	105.4	17.2	-.8	.425	.82**

Note. *N* = 20.

*$p<.05$. **$p<.01$. ***$p<.001$.

Table 1 further indicates that three of the four subtests on which significant differences appeared are on the Verbal Scale. However, the Information subtest mean IQ was significantly higher on the WISC-III, whereas the Vocabulary and Comprehension subtest mean IQs were significantly higher on the WISC-R.

The second hypothesis stated there is no significant difference in WISC-R and WISC-III indices (VC, PO, FFD).

This hypothesis was tested using a *t* test for dependent means and was tested for statistical significance. A significant difference was found for the VC index but no significant differences appeared on the PO and FFD indices. Thus the null hypothesis of no significant difference between indices was rejected for VC but not rejected for PO and FFD.

There is no significant difference in WISC-R and WISC-III IQ scale scores was the third hypothesis tested.

Using a *t* test for dependent means, this hypothesis was tested for significance. As noted on Table 1, the IQ scale scores (Verbal, Performance, and Full Scale) were significantly higher on the WISC-R; for both the Verbal and Performance Scales, significance was at the .05 level, Full Scale was at the .01 level. Thus the null hypothesis of no significant difference was rejected for all three IQ scales.

The fourth and final hypothesis stated there is no significant correlation in WISC-R and WISC-III subtest, index, and scale scores.

As noted on Table 1, correlation coefficients for the subtests reveal significant correlation (.52 to .92) on all subtests except the Comprehension and Picture Arrangement subtests (.33 and .22, respectively). Correlation coefficients for the indices revealed a significant correlation for the VC and FFD indices (.01 level) but not for the PO index. The correlation coefficients for the three IQ scales were also significant. Thus, the null hypothesis of no significant correlation was accepted for the Comprehension and Picture Arrangement subtests and rejected for all other subtests, indices, and scales.

Discussion

This study compared WISC-R and WISC-III subtest, index and scale scores of 20 rural, Caucasian, midwestern children from age 8 to 12 years old. Rater consistency was maintained by having one researcher score all the protocols. Results revealed significantly higher WISC-R Verbal, Performance, and Full Scale IQs. Of the 11 WISC-R and WISC-III subtests administered in this study, higher WISC-R scores appeared on nine of the subtests, and on the Vocabulary, Comprehension, and Block Design subtests the differences were significant. Conversely, higher WISC-III scores appeared on the Information and Digit Span subtests with the Information subtest difference reaching significance. The Digit Span scores are partially responsible for the higher WISC-III FFD index which reached statistical significance. It should be

noted in comparing the differences in means with the WISC-III Manual, the Information, Vocabulary, Comprehension, and Block Design subtest means in the present study were at least twice as great (Wechsler, 1991, p. 198). However, the Verbal Scale mean IQ score differences fell within the previously reported range of WISC-R and WISC-III differences (Wechsler, 1991, p. 197). Comparison of indices was not examined in the WISC-III Manual study.

Correlation coefficients were significantly high on all subtests except for the Comprehension subtest (Verbal) and Picture Arrangement subtest (Performance). Conversely, Comprehension correlation coefficients in the WISC-III Manual (Wechsler, 1991, p. 198) were approximately twice as high as those noted in the present study, indicating possible correlational concerns regarding the comparative Comprehension subtests.

Computed correlation coefficients for the indices indicated significant correlation on two of the three indices (VC, FFD). Significantly high correlations in the present study were obtained for the three IQ scales with the Verbal IQ scale being the lowest of the three. The correlation coefficients for the IQ scales closely coincide to those reported in the WISC-III Manual (Wechsler, 1991, p.197). Since only a portion of the subtests are used in determining the indices, reliance on these scores to evaluate the child should be approached cautiously until further research can be

done. While results of this investigation are consistent with the only published comparable study, they must be seen in the perspective of a small sample of rural, Caucasian, midwestern children.

CHAPTER 5

SUMMARY

Although the limitations of IQ scores are acknowledged, scores obtained on intelligence tests often play a significant role in the determination of the educational placement and programs deemed most appropriate for young children. Therefore, it is necessary to investigate newer versions of intelligence tests to determine if the scores obtained on both tests can be considered comparable.

With most correlation coefficients being significant in this study (.84, .54, .79 for the Verbal, Performance, and Full Scales, respectively), it is appropriate to assume that both the WISC-R and WISC-III are essentially tapping the same intellectual abilities in children. Verbal and Performance Scale mean IQ differences ranged from approximately four to seven IQ points on the WISC-R and WISC-III with the WISC-R scores being higher. The Full Scale mean IQ differences were approximately eight points with the WISC-R score again being higher. The study cited in the WISC-III Manual (Wechsler, 1991, p. 197-199) comparing group mean IQ scores of the three IQ scales of the WISC-R and WISC-III, showed correlation coefficients of .90, .80, and .86, for the Performance, Verbal and Full Scales, respectively.

Recent investigations reveal an increase in IQ scores for children over the years. Comparisons have been investigated concerning the differences in IQ scores of the

same child on the older and newer versions of the same intelligence test. Results of this study are consistent with the only other published study of this nature, in that the group mean IQ scores of the subjects are lower on the WISC-III as compared to its predecessor, the WISC-R. However, discrepancies have occurred when comparing the individual subtest results from this study with the study reported in the WISC-III Manual (Wechsler, 1991, p. 197-199). Mean group scores were obtained and two-tailed dependent *t* tests were computed on the IQ scales of the two tests. The Verbal and Performance Scales were significantly different at the .05 level, the Full Scale was significantly different at the .01 level.

It is of great importance that teachers, parents, and psychologists be aware that IQ scores on the WISC-R and WISC-III appear to be somewhat discrepant. Results of aptitude appraisal, therefore, must be qualified by evidence yielded by the available studies which suggest, consistent with other investigations, that more recent IQ tests yield lower IQs (a possible reflection of updated norms) and therefore are likely to affect the normative distribution when decisions with respect to special class placement and other diagnostic considerations are entertained. Because the WISC-III is still very new and psychologists will inevitably be changing from the WISC-R to the WISC-III, additional investigations are warranted for further study. Possible future studies may

entail a more detailed look at the low Comprehension and Picture Arrangement subtests correlations; the higher Information subtest scores obtained on the WISC-III; the importance of the VC, PO, and FFD indices in diagnosing children; and the comparability of WISC-R and WISC-III scores using a cross-sectional sample more representative of the population as a whole.

REFERENCES

Anastasi, A. (1976). Psychological testing (4th ed.). New York: Macmillan Publishing Co., Inc.

Catron, D. W., & Catron, S. S. (1977). WISC-R vs. WISC: A comparison with educable mentally retarded children. Journal of School Psychology, 15, 264-266.

Covet, T. M. (1976). Comparability of WISC and WISC-R full scale IQs for elementary school children with learning difficulties. Psychological Reports, 39, 280.

Davis, E. E. (1977). Matched pair comparison of WISC and WISC-R scores. Psychology in the Schools, 14, 161-166.

Fisher, T. J. (1992). [WISC-R & WISC-III profiles: A preliminary study]. Unpublished raw data. Cedar Falls, IA: University of Northern Iowa.

Flynn, J. R. (1984). The mean IQ of Americans: Massive IQ gains 1932-1978. Psychological Bulletin, 95, 29-51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. Psychological Bulletin, 101, 171-191.

Gironda, R. J. (1977). A comparison of WISC and WISC-R results of urban educable mentally retarded students. Psychology in the Schools, 14, 271-275.

Hamm, H., Wheeler, J., McCallum, S., Herrin, M., Hunter, D., & Catoe, C. (1976). Psychology in the Schools, 13, 4-8.

Iowa Department of Education. (1990). Rules of special education. Des Moines, IA: Author.

Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between 6 1/2 and 16 1/2 years. Journal of Consulting and Clinical Psychology, 43, 135-147.

Kaufman, A. S. (1979). WISC-R research: Implications for interpretation. School Psychology Digest, 8, 5-27.

Kaufman, A. S. (1990). Assessing adolescent and adult intelligence. Boston, MA: Allyn & Bacon.

Lippold, S., & Claiborn, J. M. (1983). Comparison of the Wechsler Adult Intelligence Scale and the Wechsler Adult Intelligence Scale-Revised. Journal of Consulting and Clinical Psychology, 51, 315-317.

Mishra, S. P., & Brown, K. H. (1983). The comparability of WAIS and WAIS-R IQs and subtest scores. Journal of Clinical Psychology, 39, 754-757.

Paal, N., Hesterly, S. O., & Wepfer, J. W. (1979). Comparability of the WISC and WISC-R. Journal of Learning Disabilities, 12, 348-351.

Quereshi, M. Y., & McIntire, D. H. (1984). The comparability of the WISC, WISC-R, and WPPSI. Journal of Clinical Psychology, 40, 1036-1043.

Sattler, J. R. (1982). Assessment of children's intelligence and special abilities (2nd ed.). San Diego, CA: Jerome Sattler.

Sattler, J. R. (1988). Assessment of children (3rd ed.). San Diego, CA: Jerome Sattler.

Schwarting, F. G. (1976). A comparison of the WISC and WISC-R. Psychology in the Schools, 13, 139-141.

Snyderman, M., & Rothman, S. (1986). Science, politics, and the IQ controversy. The Public Interest, 83, 79-97.

Wechsler, D. (1949). Manual, Wechsler Intelligence Scale for Children. New York: Psychological Corporation.

Wechsler, D. (1955). Manual, Wechsler Adult Intelligence Scale. New York: Psychological Corporation.

Wechsler, D. (1967). Manual, Wechsler Preschool and Primary Scales of Intelligence. New York: Psychological Corporation.

Wechsler, D. (1974). Manual, Wechsler Intelligence Scale for Children-Revised. New York: Psychological Corporation.

Wechsler, D. (1981). Manual, Wechsler Adult Intelligence Scale-Revised. New York: Psychological Corporation.

Wechsler, D. (1991). Manual, Wechsler Intelligence Scale for Children (3rd ed.). New York: Psychological Corporation.